

SpatialData: an open and universal data framework for spatial omics

Quentin Blampey
Luca Marconato

AI & ML in Spatial Single-Cell Transcriptomics Workshop
16 Oct 2025 - Lyon (France)



Workshop
notebooks



Documentation
notebooks

Learning resources

We believe that interoperable foundational software can accelerate science

scverse®

Foundational tools for single-cell omics data analysis

GitHub

Discourse

Zulip

Twitter

YouTube

<https://scverse.org/>

CORE PACKAGES



anndata

Standard for annotated matrices



mudata

Multimodal data format



scanpy

Single-cell analysis framework



muon

Multi-omics analysis framework



scvi-tools

Single-cell machine learning
framework



scirpy

Single-cell immune sequencing
analysis framework



squidpy

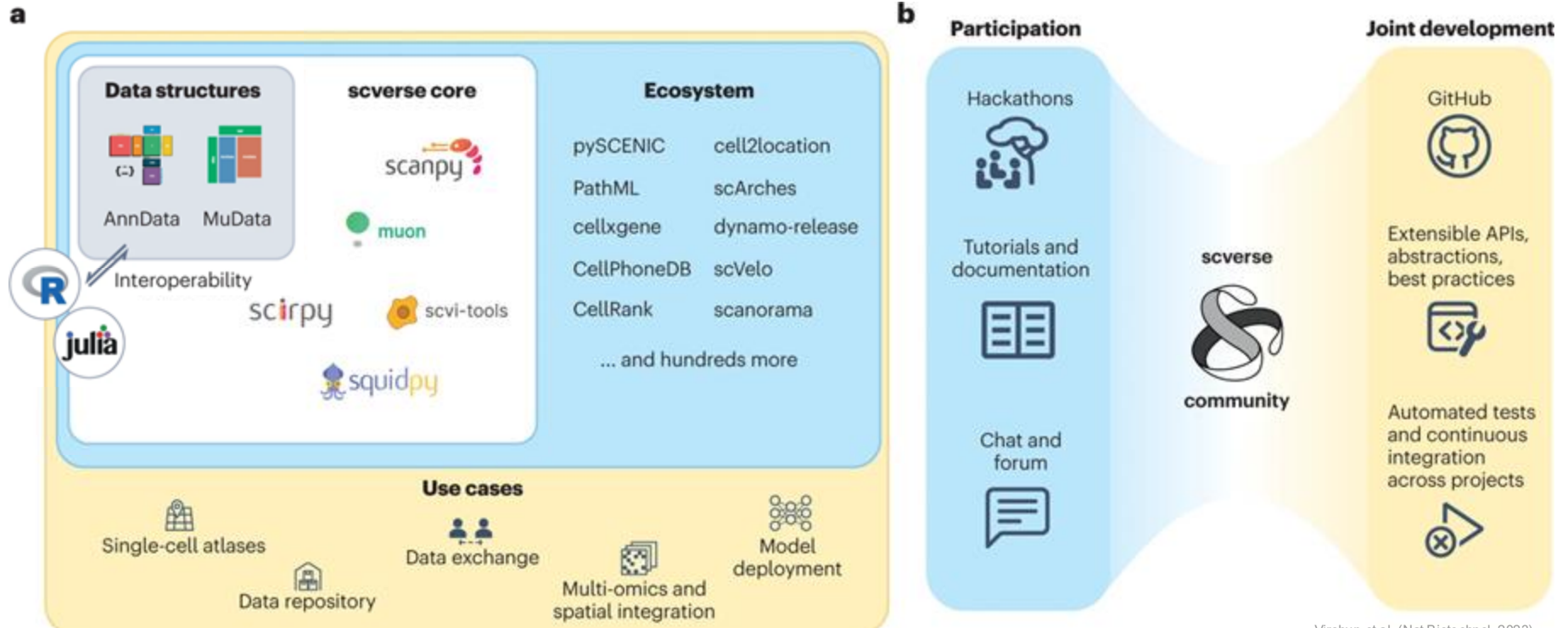
Spatial single-cell analysis



spatialdata

Spatial data format

We facilitate the growth of an ecosystem of methods, built upon our core libraries



Talks overview

From upstream packages to downstream tasks



SpatialData

Data structure for spatial omics

Official **scverse** data structure

*As a contributor, I talk on behalf
of scverse and Luca Marconato*

nature methods

This talk



Analysis toolkit for spatial
omics

scverse ecosystem package

Author and maintainer

nature
communications



Deep learning / foundation
model for spatial domains

scverse ecosystem package

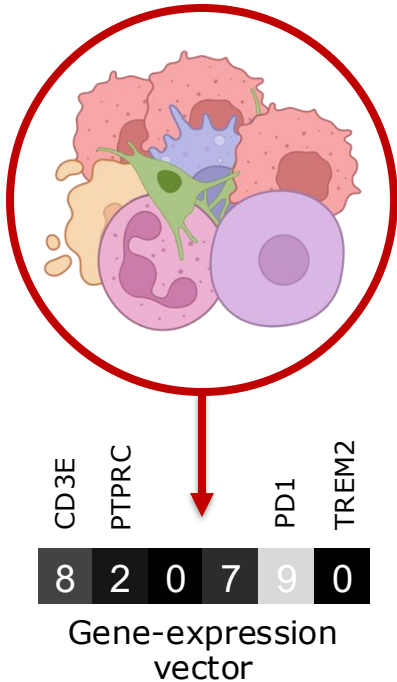
Author and maintainer

accepted in
nature methods

Spatial technologies enable the quantification of biological processes in the tissue context

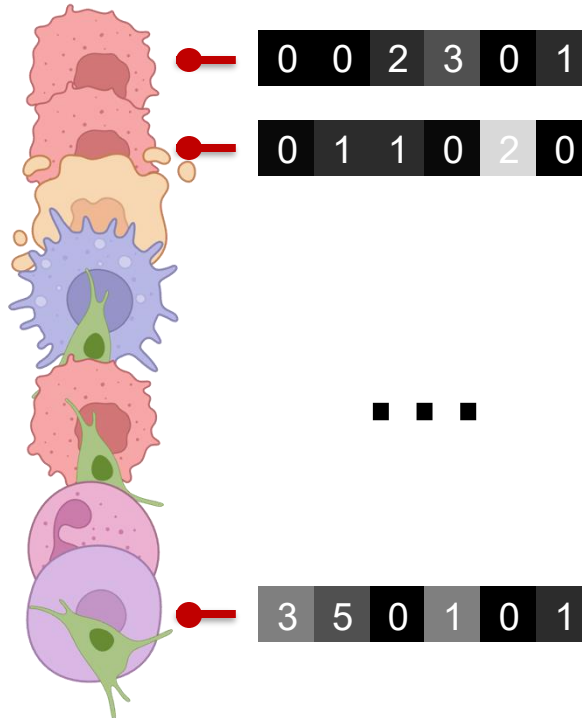
1

Bulk resolution



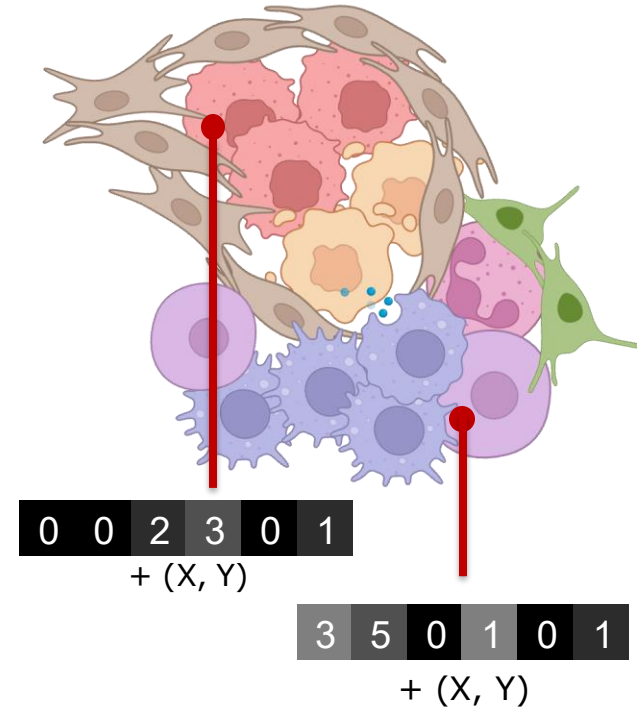
2

Single-cell resolution

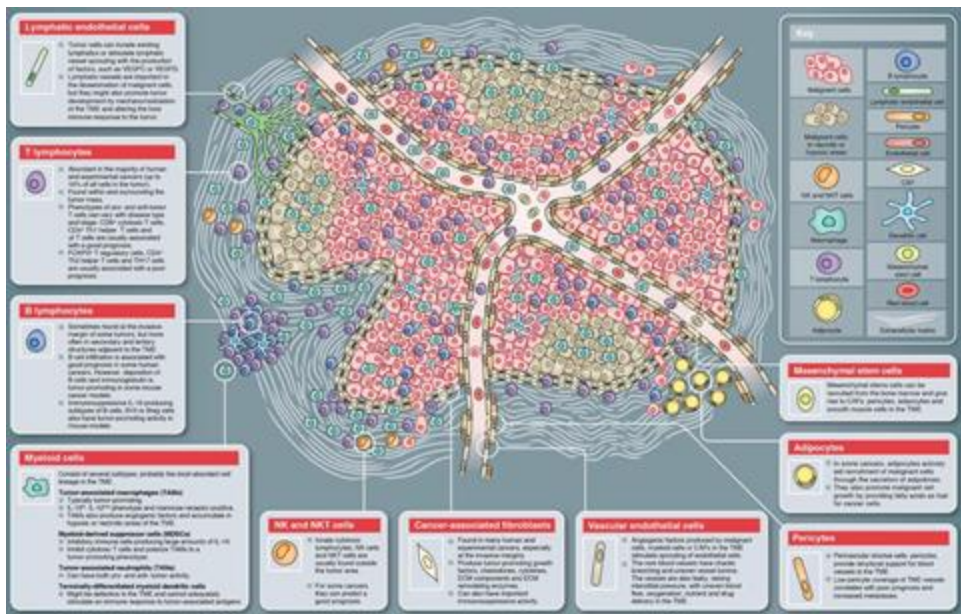


3

Spatial omics at single-cell or spot resolution

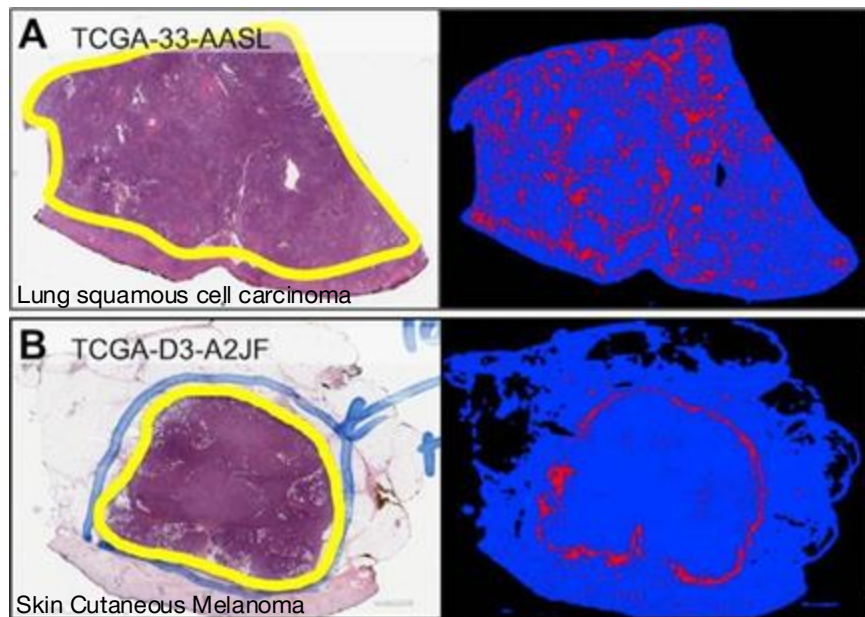


Spatial organization plays a crucial role in disease progression



Adapted from Balkwill et al. (J. Cell Sci, 2018)

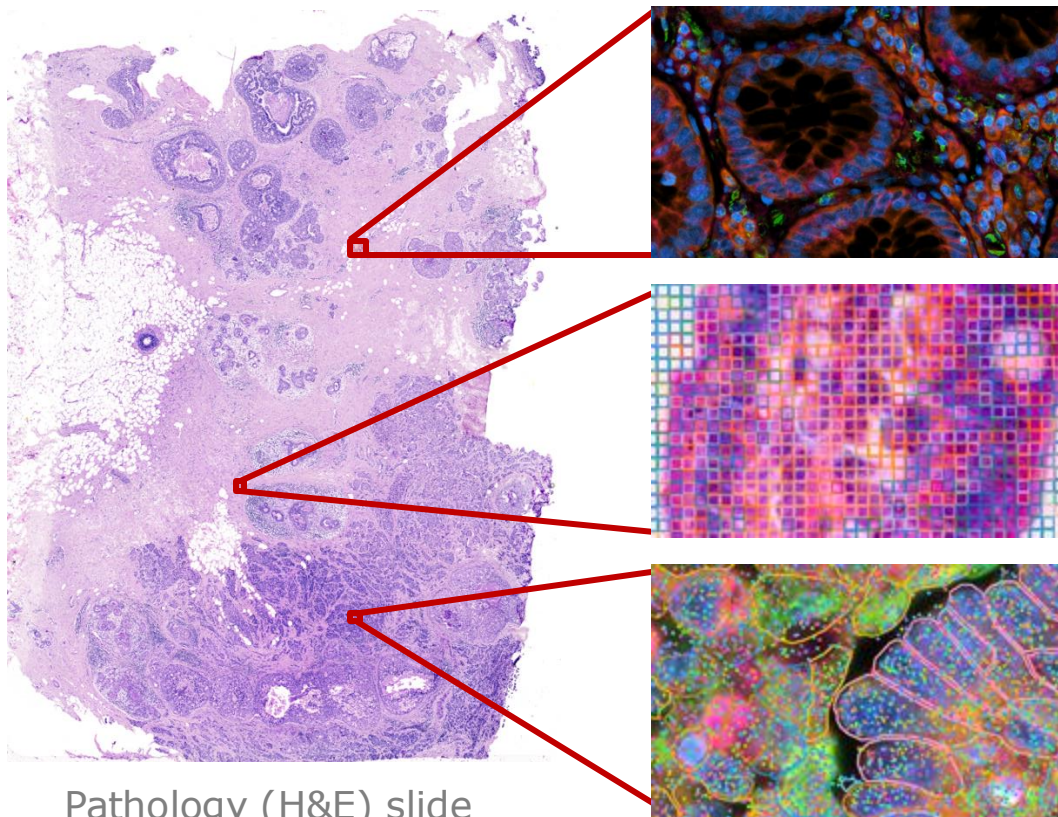
Tumor microenvironments are organized in complex spatial structures



Adapted from Saltz et al. (Cell Reports, 2018)

In this study, lymphocytes (red) infiltrating beyond the tumor boundaries (yellow) are linked with better prognosis

Different types of omics data



Pathology (H&E) slide

Spatial proteomics

Staining images (one channel = one protein)

Up to 100 proteins (3 shown here)

Spatial Transcriptomics (NGS-based)

Whole-transcriptome sequencing

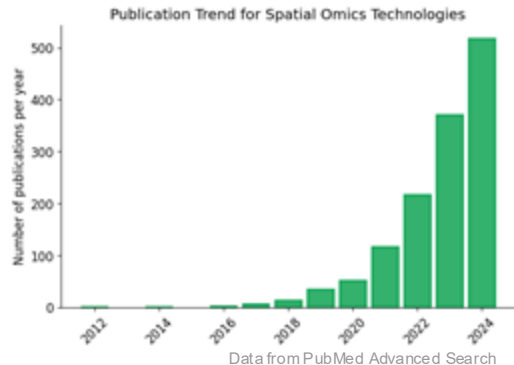
Limited to the spot resolution (subcellular resolution with the Visium HD)

Spatial Transcriptomics (imaging-based)

Subcellular resolution (3D transcript location)

Limited in number of different genes

Representing and operating on spatial omics data is becoming increasingly complex



BaristaSeq
Cartana IIS
DBiT-seq
FISSEQ
G4X
HDST
IMC/Hyperion
LCM
MALDI
MERFISH

MIBI-TOF
Molecular Cartography
NanoString CosMx SMI
NanoString GeoMx DSP
Nova-ST
osmFISH
Open-ST
PLISH
Pixel-seq
Rebus Bio

RNA SPOTs
RNAscope
SABER-FISH
seqFISH+
Seq-scope
Slide-DNA-seq
Slide-seq v2
smFISH
SpaceM
ST

Stereo-seq
StarMAP
Tomo-seq
Visium HD

Visium (SpaceRanger 1.0.0, 1.1.0, 1.2.0, 1.3.0, 1.3.1, 2.0.0, 2.0.1, 2.1.0, ...)

Xenium (Xenium Analyzer 1.0.2, 1.3.0, 1.4.0, 1.5.0, 1.6.0, 1.7.0, 2.0.0, 3.0.0, 3.0.1, ...)

XYZeq
...

Formats for raster data



Common compressed raster formats



Tensor storages



Multiscale storages

Formats for geometries



Human-readable formats



Binary formats

Formats for annotations

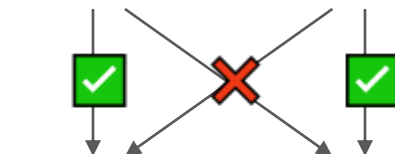


Human-readable formats



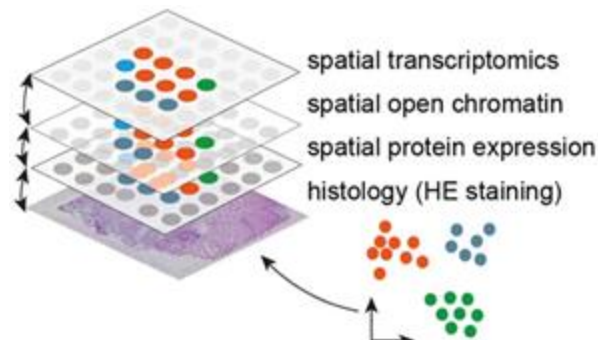
Binary formats

Technology 1 Technology 2



Software 1

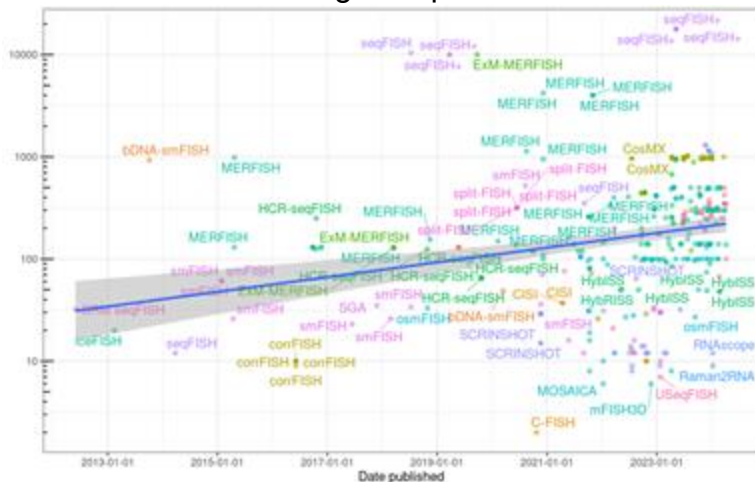
Software 2



Kiessling & Kuppe (Genome Med, 2024)

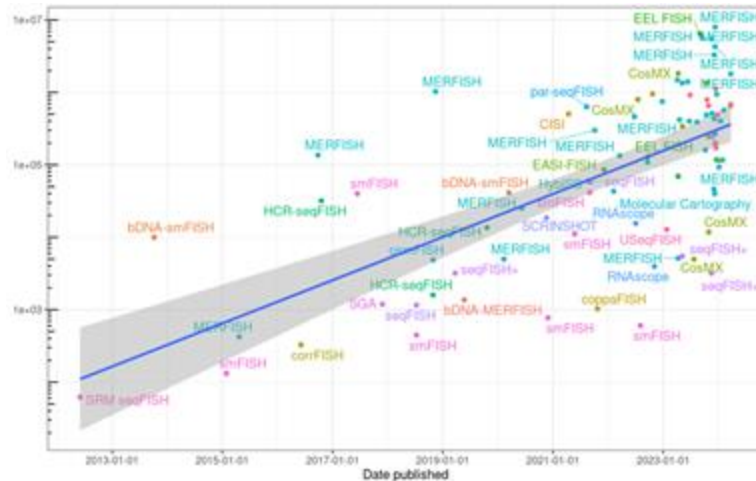
Scalability becomes pressing as datasets grow exponentially in size

Number of genes per dataset



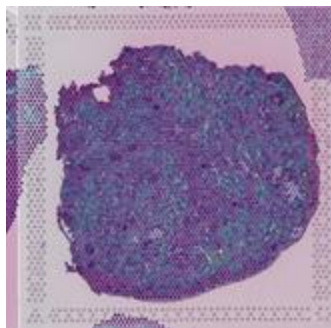
Moses & Pachter (Nat Methods, 2022)

Total number of cells per study

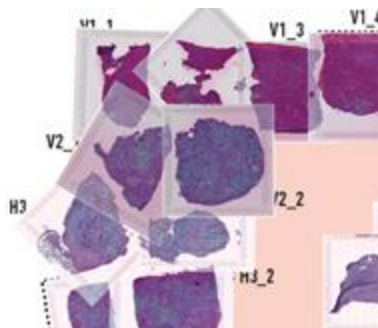


Moses & Pachter (Nat Methods, 2022)

Data can (already) be large



Single slide



Dataset from Erickson, Lundeberg et al. (Nature, 2022)

Multiple (overlapping) slides in space

To reach the goal of building a foundational infrastructure, we bridged existing communities

scverse



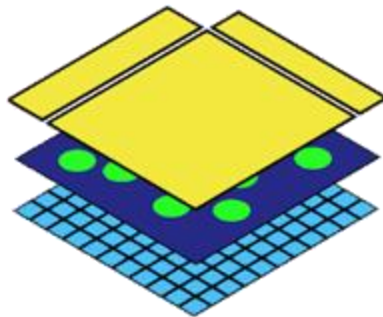
Single-cell data



AnnData



SpatialData



Marconato*, Palla*, Yamauchi*, Virshup* et al.
(Nature Methods, 2024)

Napari core



napari

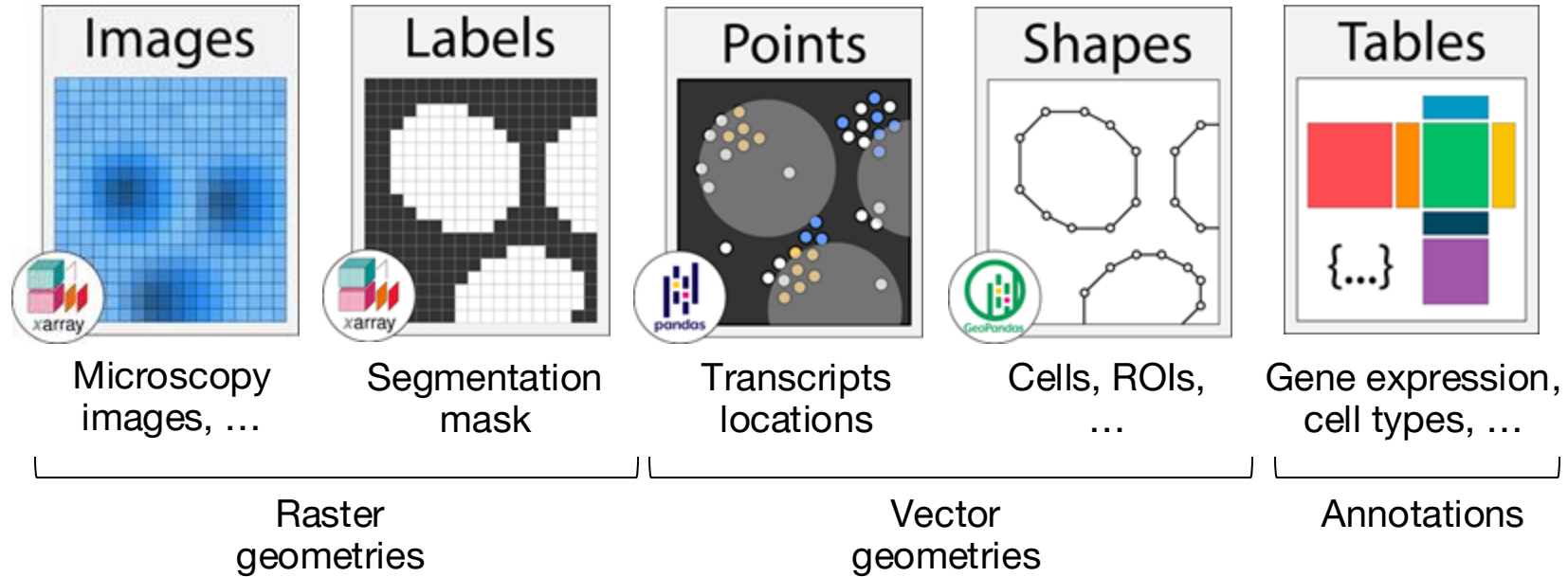
Interactive visualization

OME (Open Microscopy Environment)

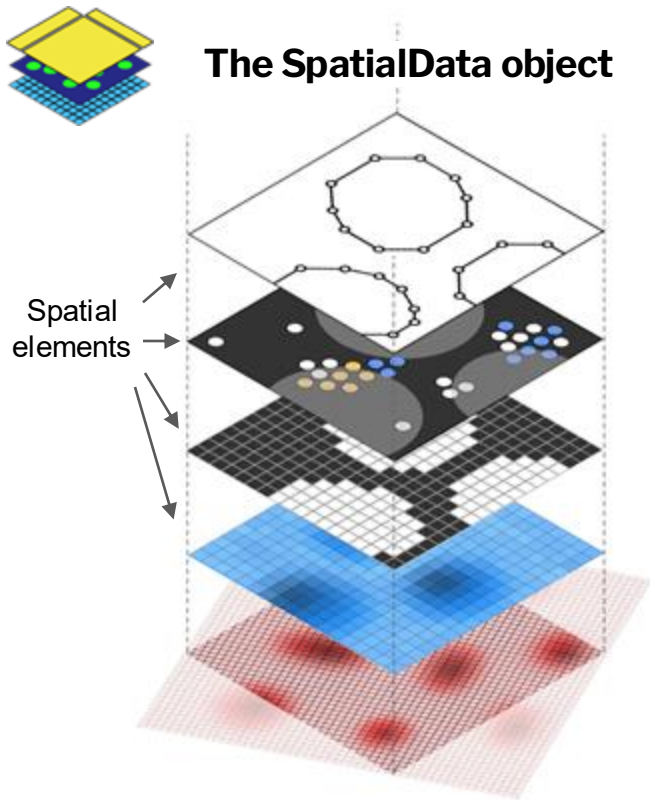


Large images, standard formats

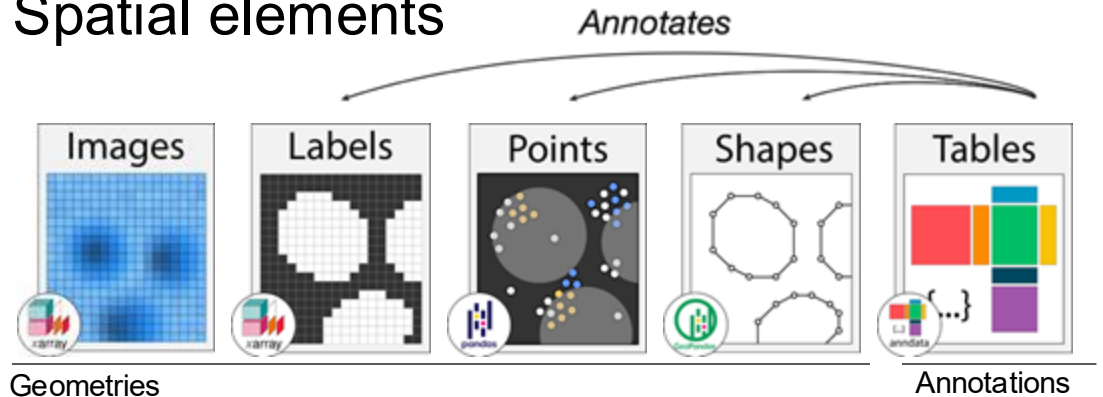
Data is represented with a modular combination of reusable elements



SpatialData introduces a modular representation for spatial omics

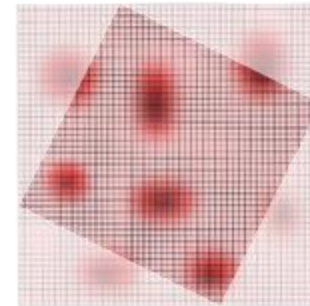


Spatial elements



Coordinate transformation

```
{
  "name": "pixels-to-micrometers",
  "type": "affine",
  "values": [[0.89, -0.45, 1.00],
            [0.45, 0.89, 2.00], [0.00, 0.00, 1.00]],
  "input_space": "",
  "output_space": "physical-micrometers"
}
```



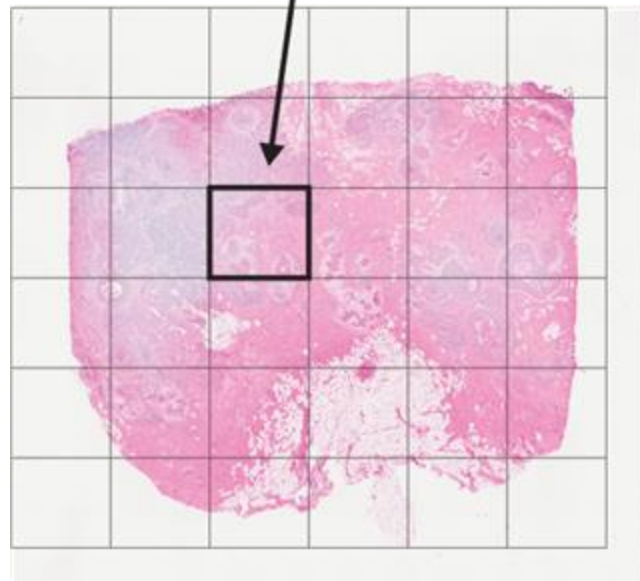
Read/write

On disk:



Chunked, multiscale representation leads to efficient image access

Zarr chunk



Scale 0
(full resolution)



Scale 1
(downscaled 500x)

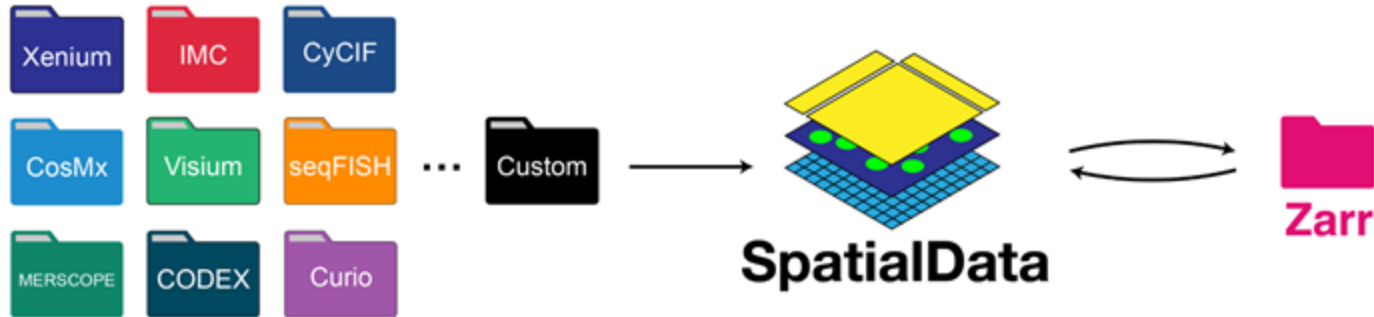


Scale 2
(downscaled 1000x)

SpatialData IO streamlines data ingestion from popular technologies

SpatialData IO (=Input Output)

```
cosmx_sdata = spatialdata.io.cosmx('/path/to/data')
```



Giovanni
Palla



Laurens
Lehner



Quentin
Blampey

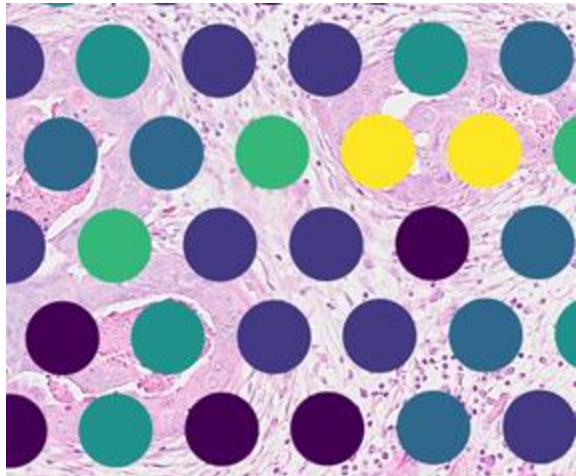


Wouter-Michiel
Vierdag

SpatialData unifies the representation of spatial omics across technologies

Data at different scales

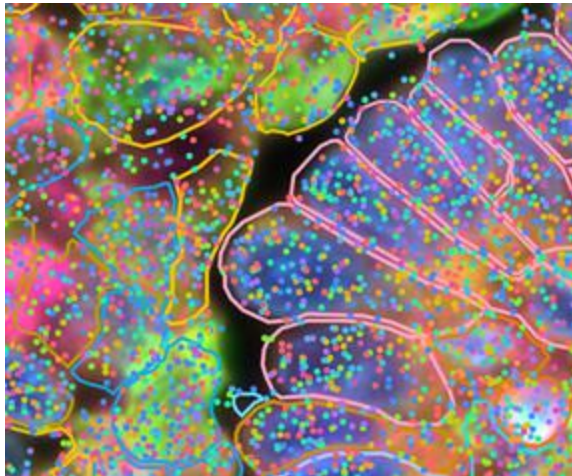
Visium



Data: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>

Resolution: 55 μ m
Transcriptome-wide

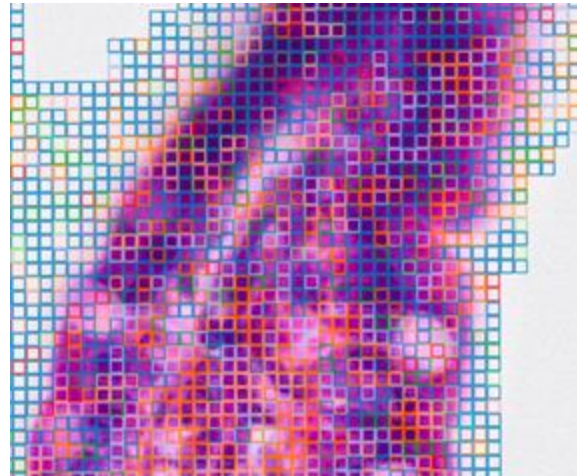
Xenium



Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation

Resolution: single-molecule
Up to 5K genes

Visium HD



Visium HD Spatial Gene Expression Library, Mouse Small Intestine (FFPE)

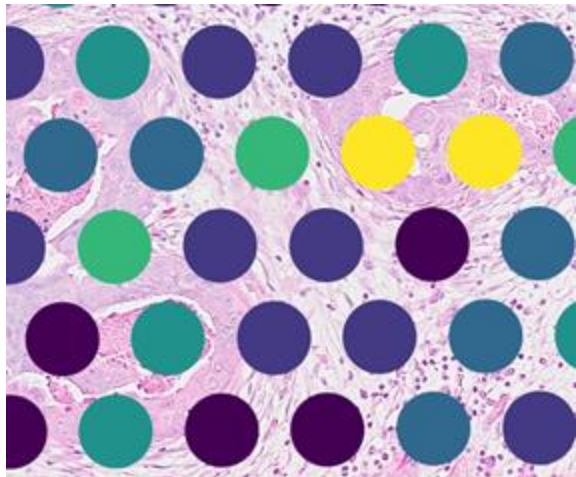
Resolution: 2 μ m, 8 μ m, 16 μ m, ...
Transcriptome-wide

The raw data is very different across technologies

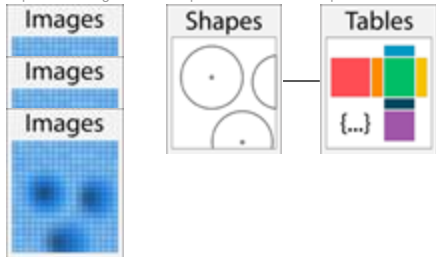
SpatialData unifies the representation of spatial omics across technologies

Data at different scales

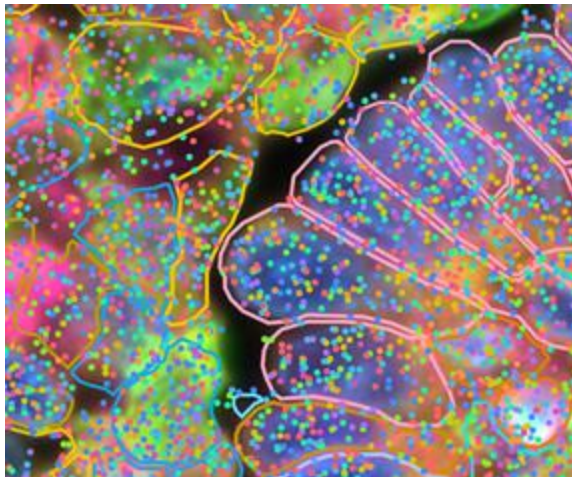
Visium



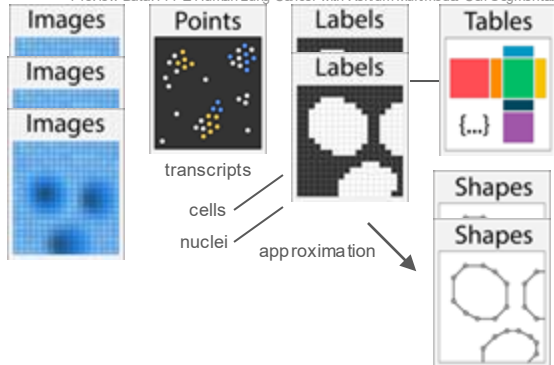
Data: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>



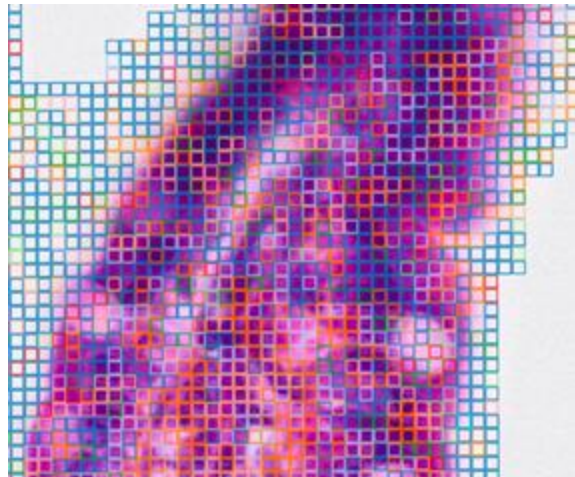
Xenium



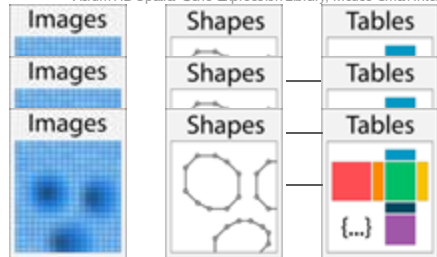
Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation



Visium HD



Visium HD Spatial Gene Expression Library, Mouse Small Intestine (FFPE)

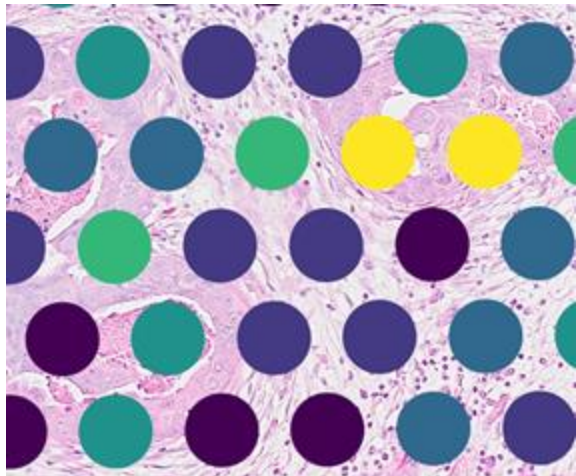


2μm, 8μm, 16μm bins

SpatialData unifies the representation of spatial omics across technologies

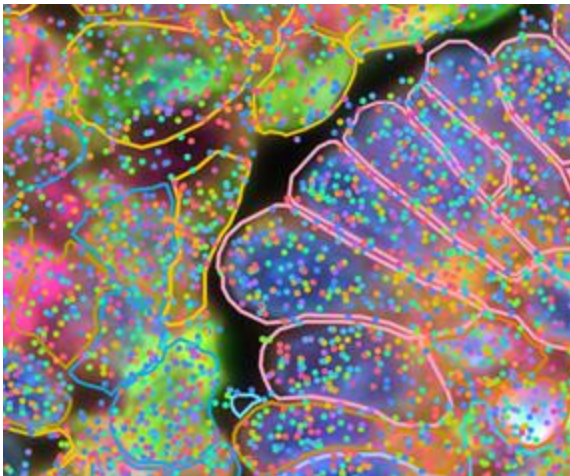
Data at different scales

Visium



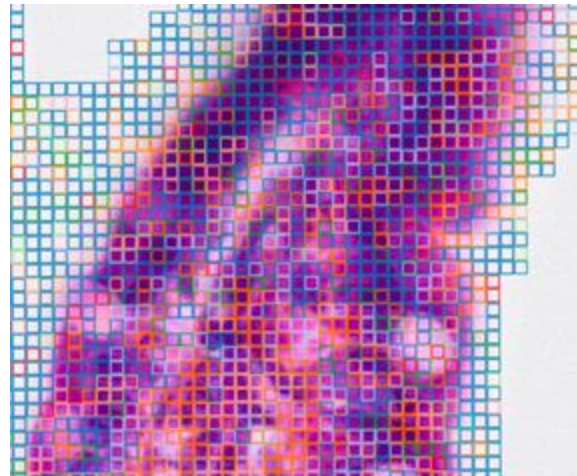
Data: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>

Xenium



Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation

Visium HD



Visium HD Spatial Gene Expression Library, Mouse Small Intestine (FFPE)

Reading the data from disk:

```
from spatialdata_io import visium  
  
sdata = visium('my_space_ranger_output')
```



```
from spatialdata_io import xenium  
  
sdata = xenium('my_xenium_analyzer_output')
```



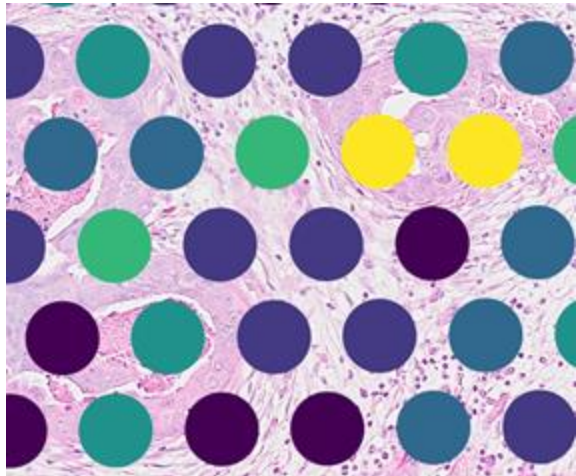
```
from spatialdata_io import visium_hd  
  
sdata = visium_hd('my_space_ranger_output')
```



SpatialData unifies the representation of spatial omics across technologies

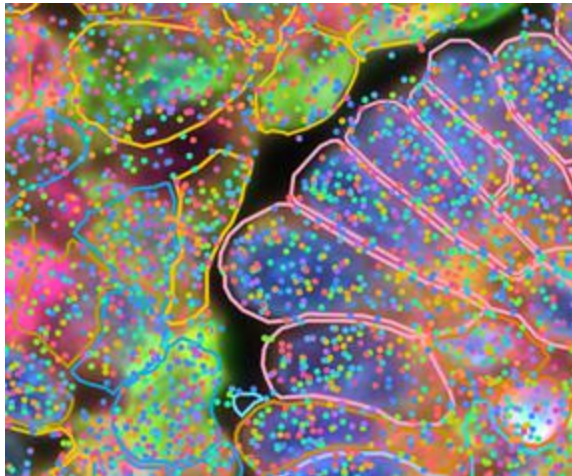
Data at different scales

Visium



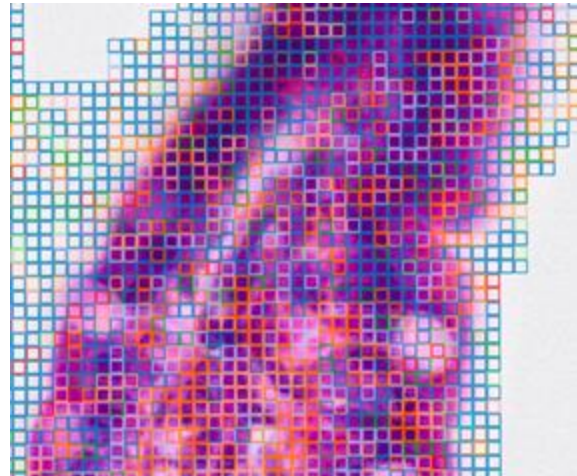
Data: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>

Xenium



Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation

Visium HD



Visium HD Spatial Gene Expression Library, Mouse Small Intestine (FFPE)



```
SpatialData object
├── Images
│   ├── 'CytAssist_FFPE_Protein_Expression_Human_Glioblastoma_hires_image': DataArray[cyx] (3, 2800, 1744)
│   └── 'CytAssist_FFPE_Protein_Expression_Human_Glioblastoma_lowres_image': DataArray[cyx] (3, 600, 523)
├── Shapes
│   └── 'CytAssist_FFPE_Protein_Expression_Human_Glioblastoma': GeoDataFrame shape: (5756, 2) (2D shapes)
├── Tables
│   └── 'table': AnnData (5756, 18085)
└── with coordinate systems:
    ├── 'downscaled_hires', with elements:
    │   └── CytAssist_FFPE_Protein_Expression_Human_Glioblastoma_hires_image (Images), CytAssist_FFPE_Protein_Expression_Human_Glioblastoma (Shapes)
    ├── 'downscaled_lowres', with elements:
    │   └── CytAssist_FFPE_Protein_Expression_Human_Glioblastoma_lowres_image (Images), CytAssist_FFPE_Protein_Expression_Human_Glioblastoma (Shapes)
    └── 'global', with elements:
        └── CytAssist_FFPE_Protein_Expression_Human_Glioblastoma (Shapes)
```

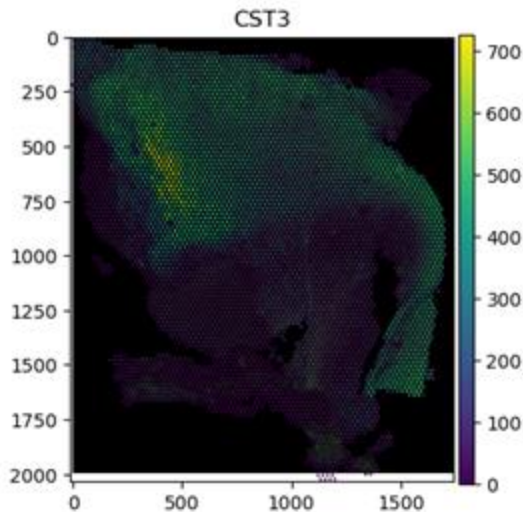
`print(sdata)`

```
SpatialData object
├── Images
│   ├── 'hm_image': DataTree[cyx] (3, 5636, 1448), (3, 2818, 724), (3, 1489, 362), (3, 784, 181), (3, 352, 90)
│   └── 'morphology_focus': DataTree[cyx] (1, 17098, 51187), (1, 8549, 25593), (1, 4274, 12796), (1, 2137, 639)
├── Labels
│   ├── 'cell_labels': DataTree[yx] (17098, 51187), (8549, 25593), (4274, 12796), (2137, 6398), (1868, 3199)
│   └── 'nucleus_labels': DataTree[yx] (17098, 51187), (8549, 25593), (4274, 12796), (2137, 6398), (1868, 3199)
├── Points
│   └── 'transcripts': DataFrame with shape: (Delayed, 11) (3D points)
├── Shapes
│   ├── 'cell_boundaries': GeoDataFrame shape: (162254, 1) (2D shapes)
│   ├── 'cell_circles': GeoDataFrame shape: (162254, 2) (2D shapes)
│   └── 'nucleus_boundaries': GeoDataFrame shape: (156628, 1) (2D shapes)
├── Tables
│   └── 'table': AnnData (162254, 377)
└── with coordinate systems:
    ├── 'global', with elements:
    │   ├── hm_image (Images), morphology_focus (Images), cell_labels (Labels), nucleus_labels (Labels), transcripts (Points), cell_boundaries (Shapes), cell_circles (Shapes), nucleus_boundaries (Shapes)
    └── 'hm_image', with elements:
        ├── hm_image (Images), morphology_focus (Images), cell_labels (Labels), nucleus_labels (Labels), transcripts (Points), cell_boundaries (Shapes), cell_circles (Shapes), nucleus_boundaries (Shapes)
```


spatialdata-plot enables static, composable visualization

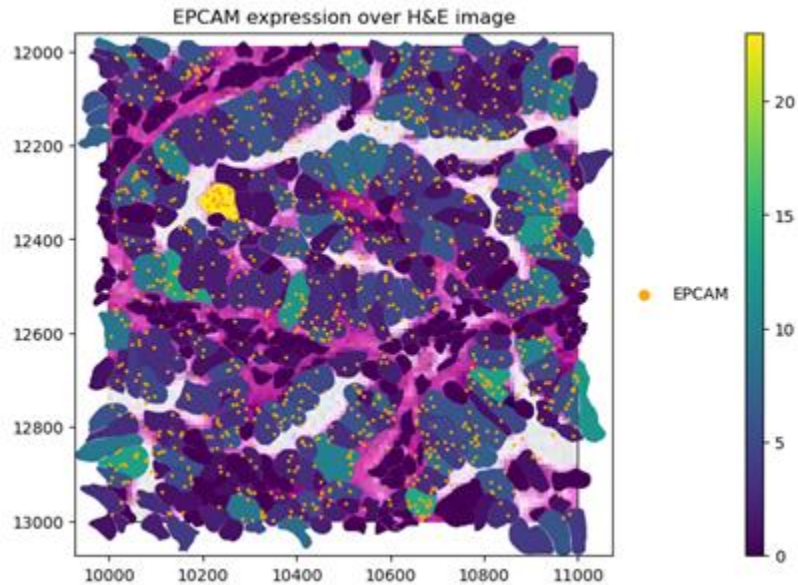
```
import spatialdata_plot

(
    sdata_visium
    .pl.render_images()
    .pl.render_shapes(color="CST3")
    .pl.show("downscaled_hires", title="CST3")
)
```



```
import spatialdata_plot

(
    sdata_xenium.
    .pl.render_images("he_image")
    .pl.render_shapes("cell_boundaries", color="EPCAM")
    .pl.render_points("transcripts", color="feature_name", groups="EPCAM", palette="orange")
    .pl.show(title=f"EPCAM expression over H&E image", coordinate_systems="global", figsize=(10, 5))
)
```



Tim Treis



Wouter-Michiel
Vierdag



Sonja
Stockhaus



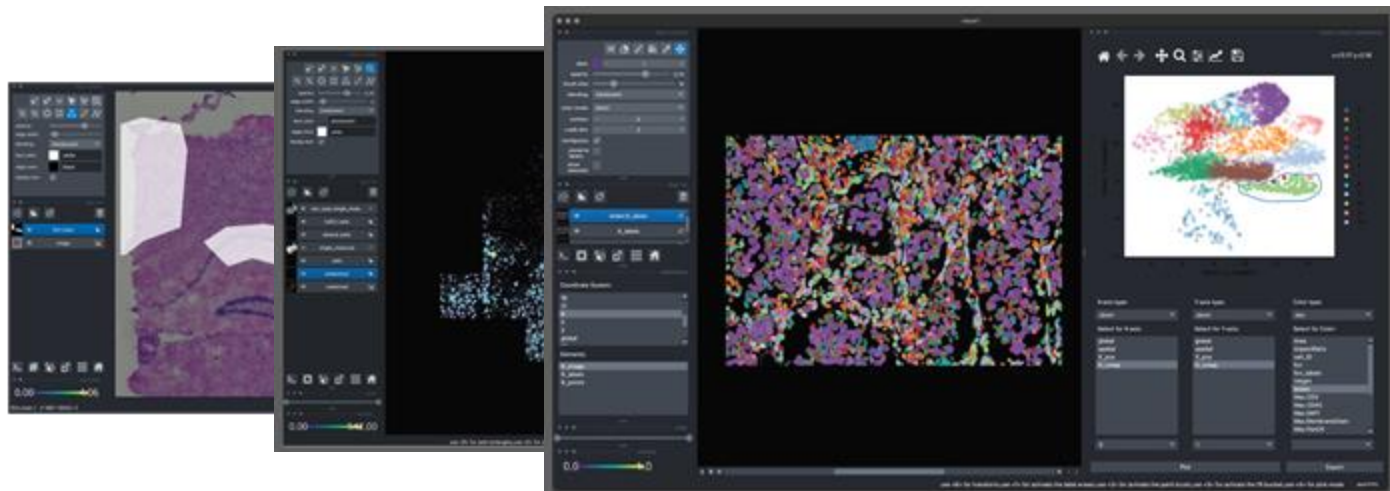
Harald
Vöringer

Napari SpatialData enables interactive data visualization and annotation

Napari plugin



```
napari_spatialdata.Interactive(cosmx_sdata)
```



Giovanni
Palla



Wouter-Michiel
Vierdag

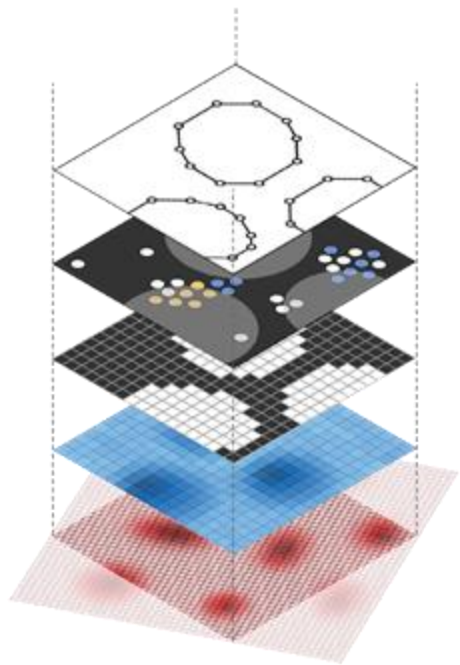


Marcela
Toth



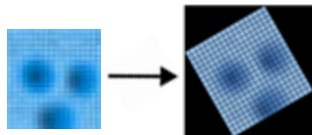
Rahul B.
Shrestha

Generalized, reusable operations are defined for SpatialData objects



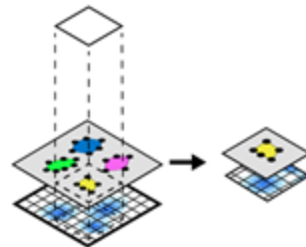
Coordinate transformations

e.g. rotate an image



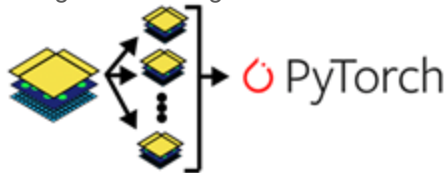
Spatial queries

e.g. crop the data



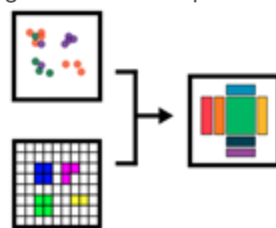
Deep learning interface

e.g. create image tiles around cells

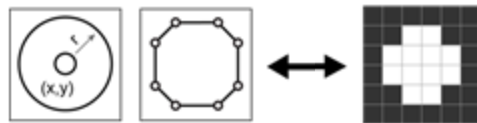


Spatial aggregations

e.g. count transcripts inside cells



Rasterize, vectorize



Documentation:

<https://spatialdata.scverse.org/en/stable/>

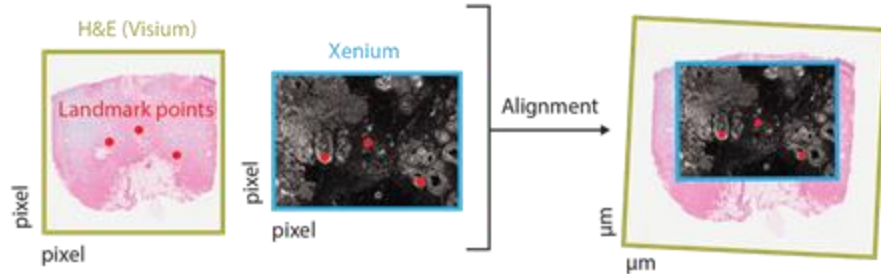
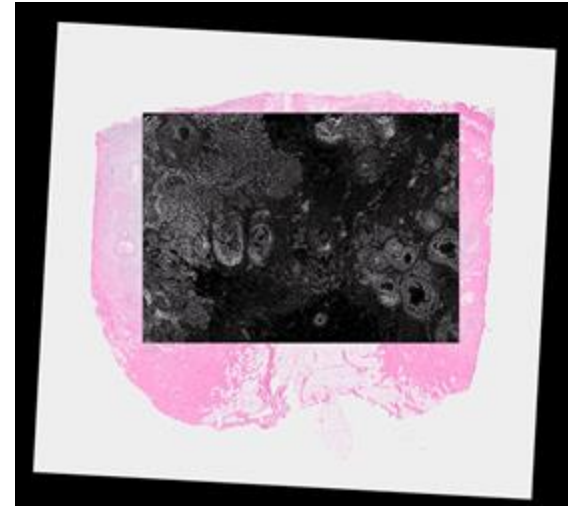
Some usage examples of these operations

Aligning multiple modalities with a landmark-based workflow

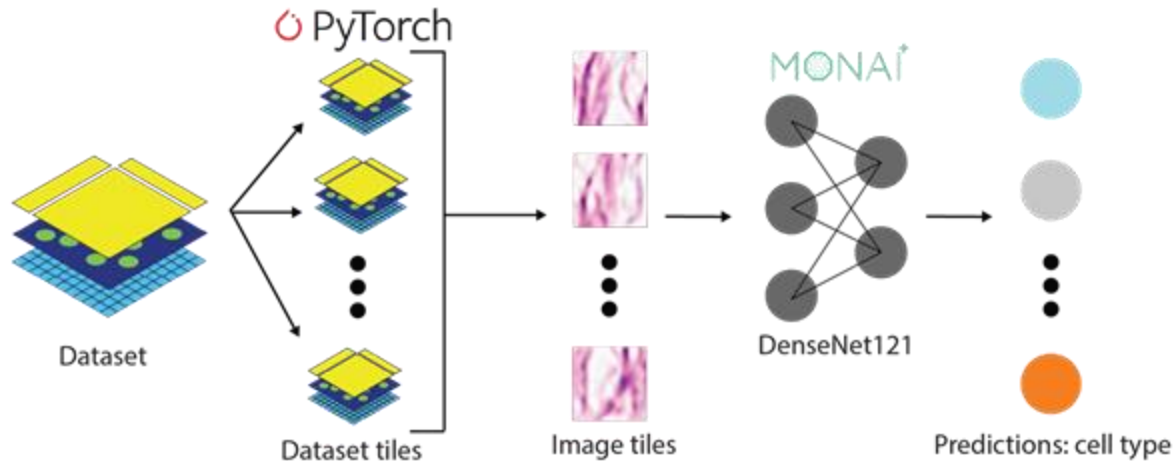
Before alignment (raw data)



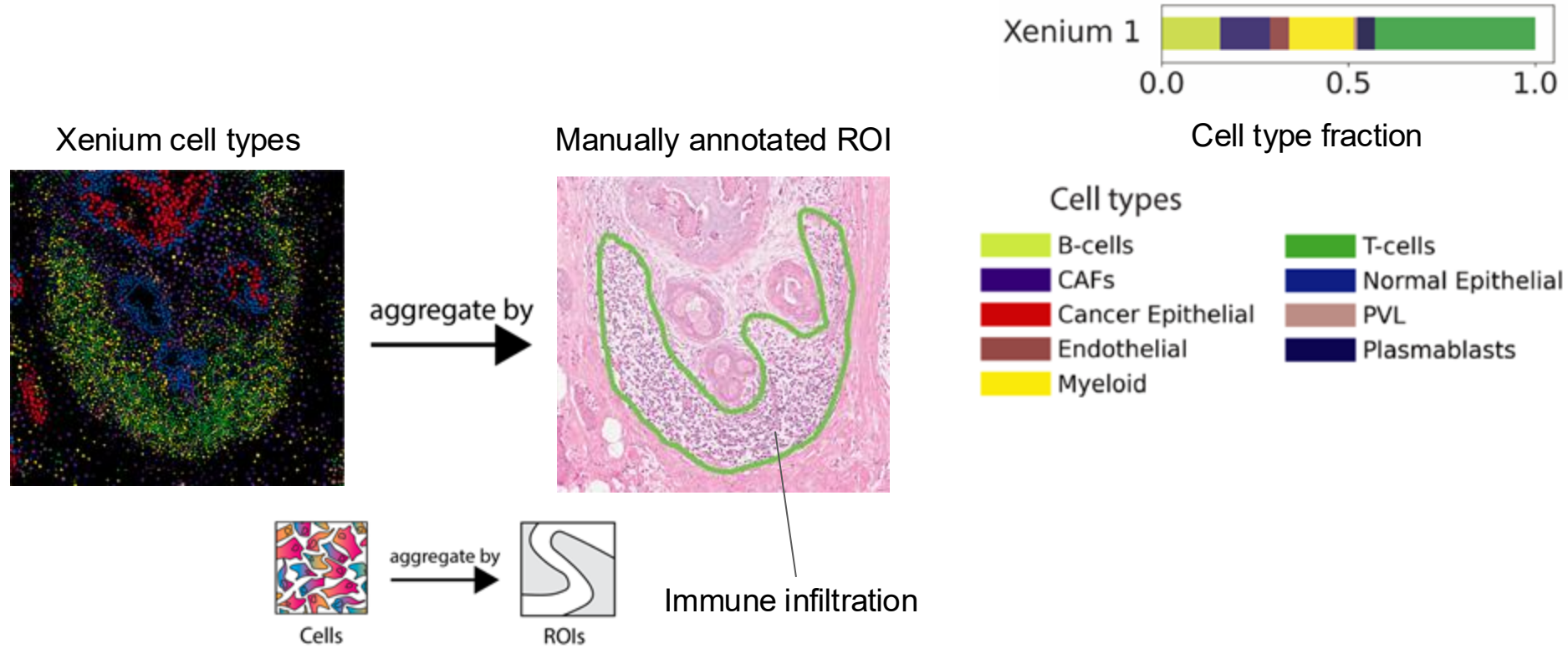
After alignment



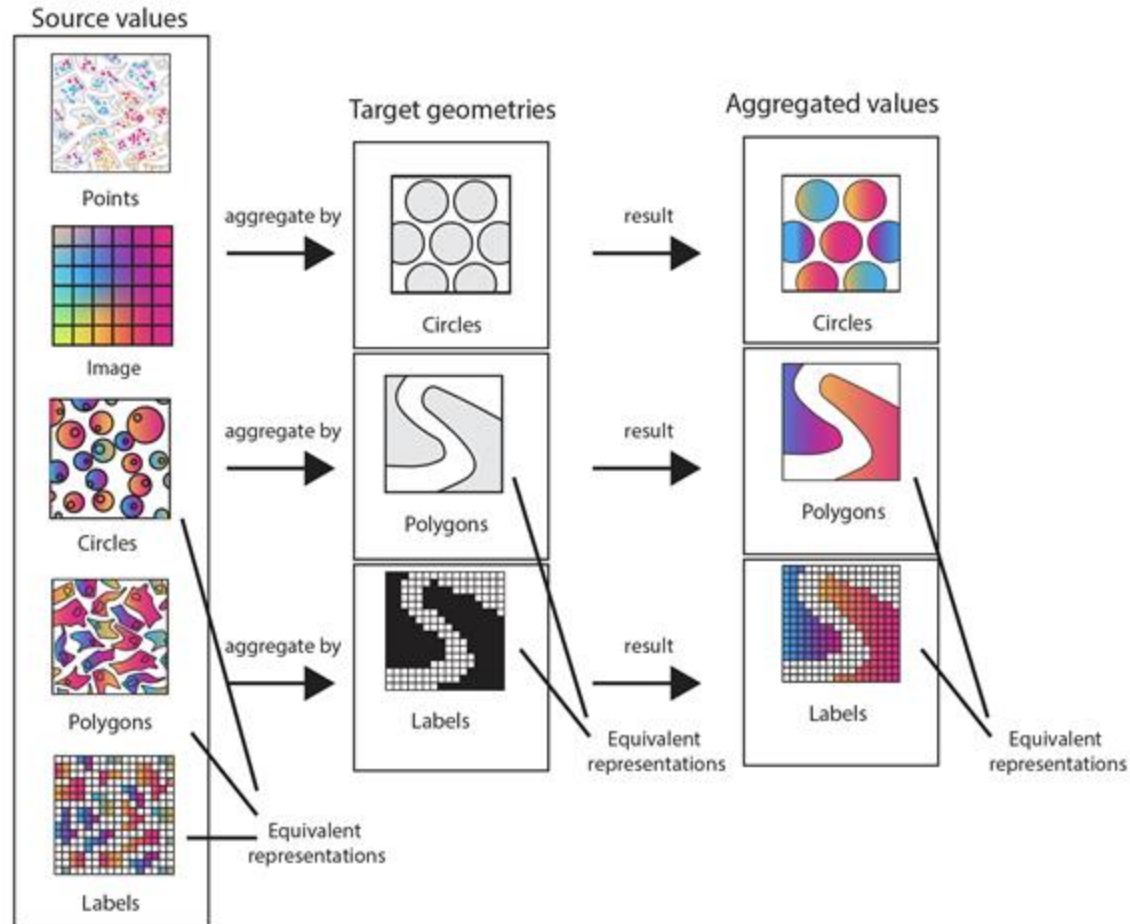
Creating deep learning datasets from multiple aligned modalities



Aggregation example: computing cell types fractions within regions of interest

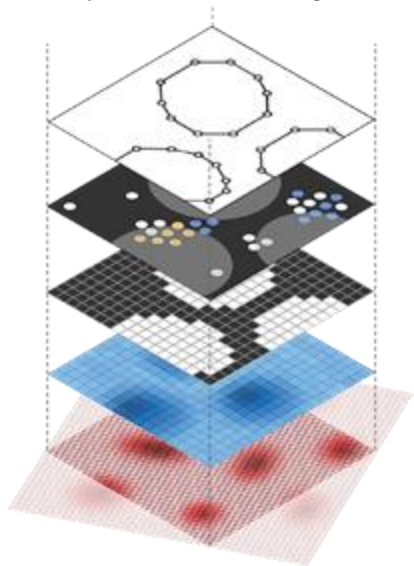


Aggregation is supported for different coordinate systems and representations



The community is enabling interoperable analysis of SpatialData objects via a growing ecosystem

SpatialData object



Dependents:

- 110 repositories
- 23 packages



spatial statistics



spatial transcriptomics
analysis



segmentation pipeline



bento-tools

subcellular analysis



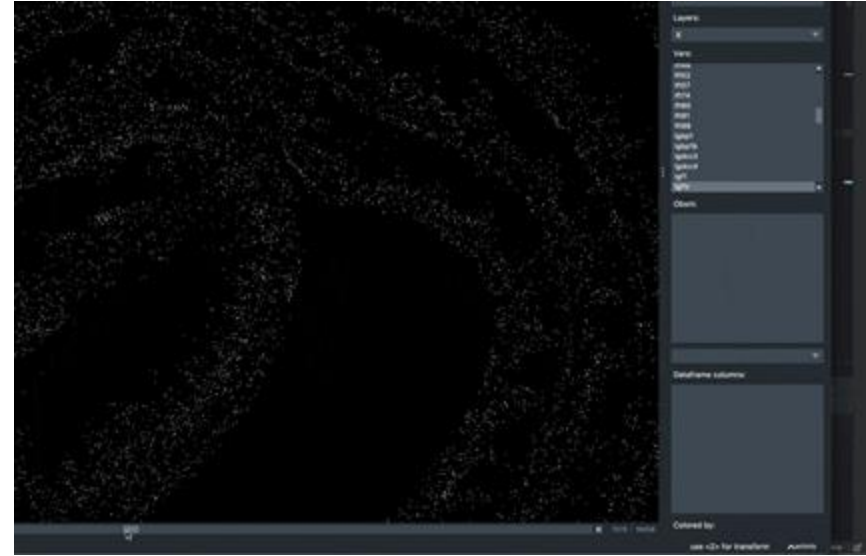
interactive visualization



spatial proteomics analysis



Very performant visualization of 2 μ m Visium HD bins in napari using rasterize_bins()



~1000x speed-up for visualizing Visium HD data in napari

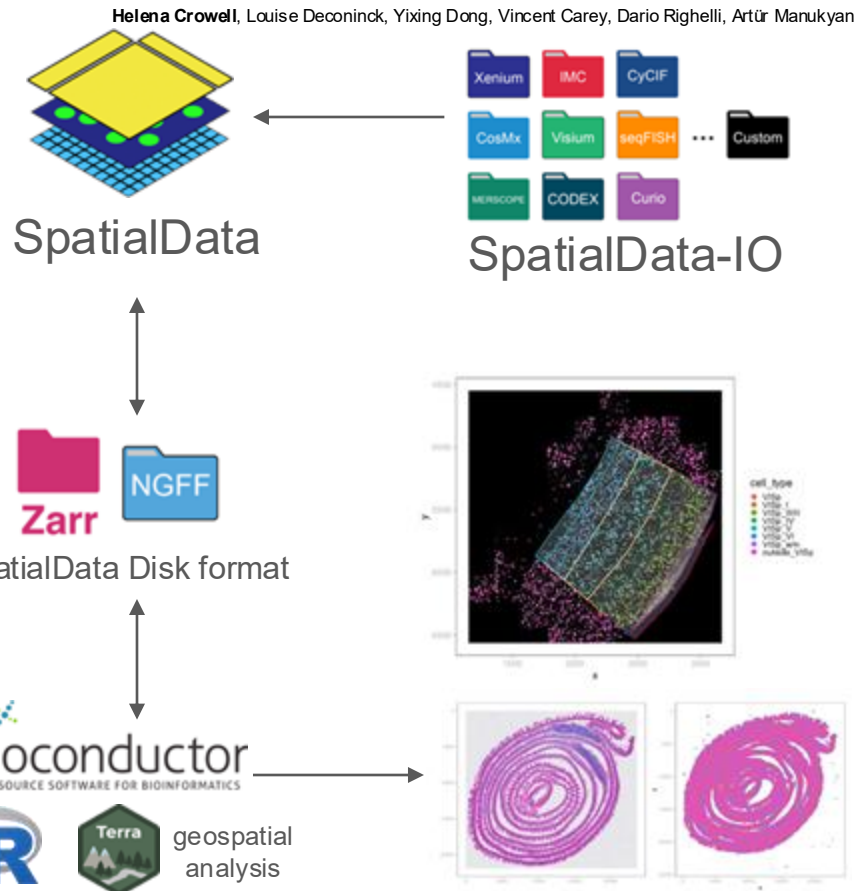
We are supporting Bioconductor developers building on top of SpatialData Zarr

work in progress!

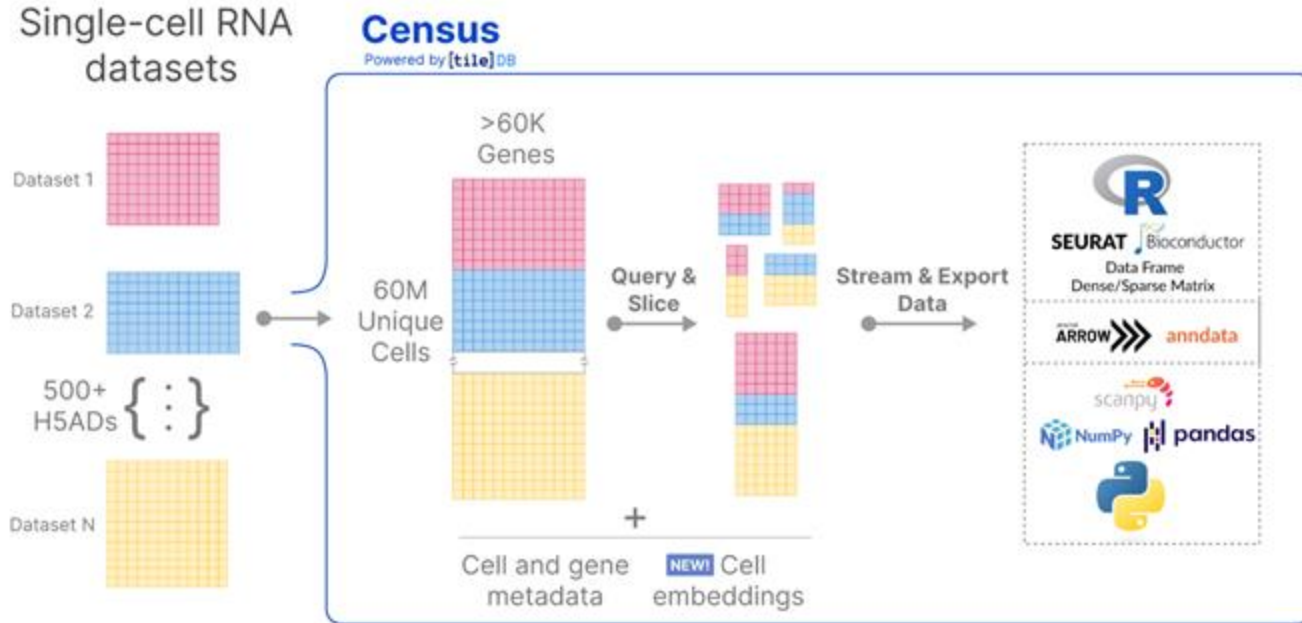
Supported by: **Chan
Zuckerberg
Initiative**



SpatialData Hackathon - Nov 2024, Basel (Switzerland)
Python & R developers working together



The CZI CELLxGENE Census database is expanding to spatial, and now exports to SpatialData



From <https://cellxgene.czi.science.com/>



We are preparing a public database of curated spatial omics datasets

🚧 work in progress! 🚧

```
# load a SpatialData dataset from the cloud
sdata_section1_query = ln.Artifact.filter(
    organisms=organisms.human,
    tissues=tissues.breast,
    description__icontains="Block A Section 1"
).one().load()
sdata_section1_query
```



example notebook
(preview)



Tim Treis



Lea
Zimmermann



Lukas
Heumos



Mark Keller

scverse > spatialdata-db

Overview Artifacts Transforms ULabels Features Projects References Bionty

Artifacts ⓘ

Datasets & models stored as files, folders, or arrays.

Advanced filtering (beta)

Artifacts Search...

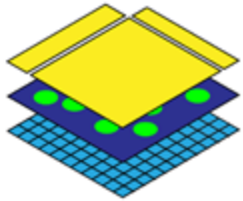
Collections

Filters

- Schema
- Created by
- ULabels
- Data format
- Bionty
 - Cell types
 - Developmental stages
 - Diseases
 - Ethnicities
 - Experimental factors
 - Organisms
 - Phenotypes
 - Tissues
- Other

Key / Description	Updated
.zarr Mouse Brain Serial (Sagittal-Posterior Section 2) · 56.2 MB ULabels Spatial Gene Expression · Specimen from Organism RNA · 2019-12-02 · Creative Commons Attribu... · Developmental stages unknown Diseases unknown Experimental factors Visium Spatial Gene Expre... Organisms mouse Tissues brain	zimea 2 weeks ago
.zarr Mouse Brain Serial · 58 MB ULabels Spatial Gene Expression · Specimen from Organism RNA · 2019-12-02 · Creative Commons Attribu... · Developmental stages unknown Diseases unknown Experimental factors Visium Spatial Gene Expre... Organisms mouse Tissues brain	upload_10X_datasets.py zimea 2 weeks ago
.zarr Mouse Brain Serial · 58.7 MB ULabels Spatial Gene Expression · Specimen from Organism RNA · 2019-12-02 · Creative Commons Attribu... · Developmental stages unknown Diseases unknown Experimental factors Visium Spatial Gene Expre... Organisms mouse Tissues brain	upload_10X_datasets.py zimea 2 weeks ago
.zarr Mouse Brain Serial · 54.8 MB ULabels Spatial Gene Expression · Specimen from Organism RNA · 2019-12-02 · Creative Commons Attribu... · Developmental stages unknown Diseases unknown Experimental factors Visium Spatial Gene Expre...	upload_10X_datasets.py zimea 2 weeks ago

Conclusions and acknowledgements



The SpatialData framework:

- **data standard** for spatial omics
- **spatial operations** (aggregation, ROI selection, rasterization, ...)
- current work on **robustness, scalability**
- **Not an analysis library**



Luca Marconato



Giovanni Palla



Kevin Yamauchi



Isaac Virshup



Tim Treis



Wouter-Michiel Vierdag



Josh Moore



Sonja Stockhaus



Elyas Heidari



Marcela Toth



Quentin Blampey



Laurens Lehner



Andreas Eisenbarth



Rahul B. Shrestha



Benjamin Rombaut



Kasia Kedziora



Lotte Pollaris



Harald Vöringer



Oliver Stegle



Fabian Theis



Moritz Gerstung



Sinem Saka



Yvan Saeys



Wolfgang Huber



Nature Methods, 2024

First authors are underlined

EMBL

Oliver Stegle
Luca Marconato
 Sinem Saka
 Wouter-Michiel Vierdag
 Wolfgang Huber
 Harald Vöhringer
 Constantin Ahlmann-Eltze
 Mike Smith

Helmholtz Munich

Fabian Theis
Giovanni Palla
Isaac Virshup
 Tim Treis
 Sonja Stockhaus
 Laurens Lehner
 Marcella Toth
 Rahul Shrestha

DKFZ

Ilia Kats
 Tobias Graf
 Moritz Gerstung
 Elyas Heidari

Josh Moore (OME, GerBi)
Kevin Yamauchi (ETH)
 Yvan Saeys (UGhent)

Lotte Pollaris (UGhent)
Benjamin Rombaut (UGhent)
 Grzegorz Bokota (UW)
 Tong Li (Sanger)
 Christian Tischer (EMBL)
 Andreas S. Eisenbarth (EMBL)
 Omer Bayraktar (Sanger)
 Ilan Gold (Helmholtz)

Helena Crowell (CNAG)
 Nils Eiling (UZH)
 Will Moore (OME, U Dundee)
 Quentin Blampey (UParis Saclay)
 Florian Wünnemann (UKHD)
 Mark Keller (HMS)
 10x Genomics team
 CZI cellxgene team
 ...

Funded by



spatialdata.scverse.org

